

MicroBent: Microbial activity at the Bentonite barrier in a deep geological nuclear waste repository

Data Management Plan

Project Name	Microbial activity at the Bentonite barrier in a deep geological nuclear waste repository
Project Acronym (number)	MicroBent (101060228)
Funding	European Commission HORIZON-MSCA-2021-PF-01
Coordinator	Margarita López-Fernández
Beneficiary	University of Granada, Spain
Deliverable Number	D1.1
Deliverable Title	Data Management Plan
Issue Date	05/09/2023
Status	Submitted
Dissemination level	Public
Contact	margaritalopez@ugr.es
Version	1.0

Version	Modifications	Date	Author
1.0	Initial document	05/09/2023	Margarita López-Fernández

MicroBent - Data Management Plan

1. DATA SUMMARY

1.1. Introduction and purpose

This is the initial **Data Management Plan** (DMP) for the project **MicroBent (101060228)**, funded by the European Commission through the Marie Skłodowska-Curie Actions (MSCA) as part of the Postdoctoral Fellowships 2021 (HORIZON-MSCA-2021-PF-01). The coordinator of this DMP, who is also the experienced researcher associated to the funded fellowship, is **Dr. Margarita López-Fernández** from the Department of Microbiology at the **University of Granada** (UGR), Spain (the beneficiary institution). The purpose of this DMP is to provide a detailed description of the procedures and protocols for the management of the datasets generated during the lifetime of the project.

This DMP will describe the main data management principles in terms of data standards and metadata, sharing, archiving, preservation, and security.

This is an alive document that will be updated at regular intervals during the lifetime of the project and be allocated in the institutional repository of the **UGR (DIGIBUG)** under the name of **MicroBent_DMP_V1.0.pdf**

1.2. Relevance to project targets

MicroBent will generate several datasets of different types both quantitative and qualitative. The data management will serve to support the project scientific objectives and spread the project results. This includes the management of three main data categories:

1- Research and Innovation objectives.

The datasets associated to this category will allow any potential user to replicate the main scientific results of the project.

2- Dissemination activities for expert audiences.

The datasets associated to this category will give access to any potential user to the documents summarizing the main scientific results of the project. This includes preprints, technical reports (*e.g.* protocols), and conference presentations.

3- Communication activities for non-experts audiences.

The datasets associated to this category will give access to any potential user to the documents dedicated to educational purposes. This includes presentations used in events for distinct non-expert audiences and teaching material for undergraduate and graduated students.

2. DATA SET DESCRIPTIONS

MicroBent aims to investigate the indigenous microorganisms under *in situ* repository conditions, to ensure the safety of Deep Geological Repositories (DGRs) of nuclear waste. To this purpose, MicroBent proposes a multidisciplinary approach, including geology, chemistry and molecular biology to describe microbial populations and metabolisms in a repository system, and directly applied to the storage of highly

radioactive waste. Apart from the research and innovation objectives, MicroBent will generate data to disseminate its results between different expert audiences and data for educational purposes.

MicroBent will generate data with different formats that will be accessible using free softwares. The different formats are presented in the table below:

Data Type	Description	Format
Raw numerical data	Experimental data	dat
Raw sequences data	Experimental data	fastq
Images	Scientific and educational purposes	JPEG, TIFF, PDF
Reports	Scientific, technical, and educational purposes	PDF
Paper preprints	Green open access according to HORIZON EUROPE guidelines	PDF
Presentations	Scientific and educational purposes	PDF/ mp4
Compressed data	Data storage	TAR.GZ

MicroBent will create data from three distinct origins: experimental, paper preprints, and presentations (technical and pedagogical). The following tables summarize the different dataset types that will be generated by MicroBent linked to its respective work package. The total estimated amount of data is around 11GB, with datasets ranging from 10MB to 5000MB.

Work Package 1. Management			
Dataset type	Origin	Format	Quantity
Data Management Plan (DMP)	Report	PDF	< 10 MB
Coordination & Evaluation seminars	Presentation	PDF	< 300 MB
REA reports	Report	PDF	< 10 MB

WP2: Training			
Dataset type	Origin	Format	Quantity
Career Development Plan (CDP)	Report	PDF	< 10 MB
Participation in training courses, seminars, and workshops	Presentation	PDF	< 300 MB

WP3: Bentonite preparation, geochemical and mineralogical characterization			
Dataset type	Origin	Format	Quantity
X-Ray Diffraction (XRD) measurements	Experimental	PDF	< 100 MB
HAADF-STEM images	Experimental	JPEG, TIFF, PDF	< 500 MB
HRTEM-EDX images	Experimental	PEG, TIFF, PDF	< 500 MB
ICP-MS measurements	Experimental	PDF	< 100 MB
X-Ray Fluorescence (XRF) measurements	Experimental	PDF	< 100 MB

WP4: Microbial diversity			
Dataset type	Origin	Format	Quantity
DNA measurements	Experimental	PDF	< 100 MB
DNA images	Experimental	JPEG, TIFF, PDF	< 500 MB
Sequencing data	Experimental	fastq	< 1000 MB

WP5: Microbial activity			
Dataset type	Origin	Format	Quantity
RNA measurements	Experimental	PDF	< 100 MB
RNA images	Experimental	JPEG, TIFF, PDF	< 500 MB
Sequencing data	Experimental	fastq	< 5000 MB

WP6: WP6: Dissemination, Exploitation and Communication Outreach activities			
Dataset type	Origin	Format	Quantity
Dissemination, Exploitation, Communication and Outreach activities Plan	Report	PDF	< 10 MB
Conference presentations and posters	Experimental	JPEG, TIFF, PDF	< 500 MB
Manuscripts	Paper preprint	PDF	< 500 MB
Teaching material	Presentation	PDF	< 300 MB
Outreach activities	Presentation	PDF	< 300 MB
Press releases, website posts and social networking	Reports/presentations	PDF/ TIFF	< 300 MB

3. FAIR DATA

3.1. Making data Findable

To make data accessible, visible, and re-usable, the metadata system used for the description of the materials hosted in the UGR repository (DIGIBUG), is Dublin Core Qualified. This is a metadata initiative adopted by the European repository OpenAIRE. DIGIBUG assigns a unique identifier (handle) to each document and/or dataset, which allows the identification and citation of electronic documents.

Each dataset generated by MicroBent will be recorded by a dataset identifier in the general file MicroBent_Metadata_List.pdf, which will be hosted (and regularly updated) in DIGIBUG. The information of each individual dataset will be included in a linked metadata file and will contain the information appearing in the table below.

Contents of a generic metadata file associated to a given dataset	
Data identifier	The ID will result from the naming convention provided in the next table
Dataset title	The title of the dataset, which will be easily searchable and findable

Responsible partner	Partner institution responsible for the creation of the dataset
Work package	Project work package
Dataset description	A brief description of the dataset
Dataset benefit	What are the benefits of the dataset
Dataset dissemination	Where will the dataset be disseminated
Format type	See table with formats in section 2
Expected size	Dataset size (see work package tables in section 2)
Source	How was the dataset generated
Repository	DIGIBUG (arXiv also for final preprints, see section 4)
DOI (if known)	The DOI will be entered once the dataset has been deposited
Date of Submission	The date of submission will be added once the dataset has been uploaded to the repository
Keywords	Keywords associated with the dataset
Version Number	Version number to keep track of changes

The dataset identifier (first field in the previous table) will be created according to the convention presented in the table below. The corresponding metadata file will be named by adding META at the end of the identifier of its linked dataset file. All the metadata files will have extension .dat.

Convention for creating the dataset identifier	
Components	Example
Project name	MicroBent (always)
Two digits chronological number (corresponding to the appearance order in MicroBent_Metadata_List.pdf)	01
Title of the dataset	DNA sequences
Version of the dataset (DIGIBUG keeps several versions)	V1.0
Work Package associated to the dataset WP1	WP1
Format of the dataset	TAR.GZ
Example dataset file identifier: <i>MicroBent_01_DNA_sequences_V1.0_WP1.TAR.GZ</i> Example metadata file identifier: <i>MicroBent_01_DNA_sequences_V1.0_WP1_META.dat</i>	

3.2. Making data openly accessible

The data hosted in the institutional repository DIGIBUG will be accessible to the complete research community. To this respect, the data generated by MicroBent does not entail Intellectual Property (IP) rights such as patents, trademarks, or copyrights. The agreement, supervised by UGR's Research Results Transfer Office, between the beneficiary institution (UGR) and the DMP coordinator complies with the IP-MSCA rules for access rights (royalty-free basis) and results ownership. The data generated will be accessible through the DIGIBUG website and will be open to any user without restrictions.

The table below provides examples of software packages to open and/or use the data generated by MicroBent for different operating systems (Windows and MacOS) and for all the formats. These examples are not unique, and the potential user might use other standard packages and/or platforms.

Format	Examples of software (operating system) to open/use them
TAR.GZ	7-Zip (Windows), tar utility (MacOS)
dat	Notepad (Windows), GNU emacs (MacOS)
PDF	Adobe Reader (Windows), Preview (MacOS)
JPEG, PNG, TIFF	Adobe Illustrator (Windows), Preview (MacOS)
mp4	VLC (Windows), MPlayerX (MacOS)

3.3. Making data interoperable

MicroBent aims to collect and document all the data generated in a standardized way (Dublin Core Qualified) to ensure that all datasets, which will be accompanied by the corresponding metadata file, can be interpreted and shared.

Therefore, a metadata file will be created and linked to each dataset. These metadata files will include all the information detailed in the first table of section 3.1 (Contents of a generic Metadata file associated to a given Dataset).

3.4. Increase data re-use

The datasets will be made available for their re-use and be stored in DIGIBUG without any cost. If datasets are updated, the coordinator of this DMP will be responsible for managing the different data versions, making sure that the latest version is available. The policies adopted by DIGIBUG concerning licenses, availability periods, and quality are:

1. License: All materials published in the DIGIBUG repository incorporate different licenses of the nonprofit organization Creative Commons, in particular, the Creative Commons 4.0 version of the Non-Commercial-ShareAlike CC BY-NC-SA license, which is the one recommended by OpenAIRE. This license allows the reuse of data at the end of the project and the use by third parties.
2. Availability: The data will remain reusable after the end of the project with no time limitations nor access restrictions, unless embargo or access restrictions are eventually indicated.
3. Quality: The quality of the datasets is guaranteed by the DIGIBUG operating software, which performs routine backups and checking of the material hosted.

4. RESOURCES ALLOCATION

All the datasets generated by MicroBent will be allocated in DIGIBUG without costs (neither in the short nor in the long term), time limitation, or access restrictions. In addition, another version of the produced preprints will be uploaded on the arXiv (a private not-for-profit educational repository owned and operated by Cornell University). The same preprint title as the one used for the version allocated in DIGIBUG will be used

for the arXiv version. With this we will ensure and enhance green open access according to the Horizon Europe guidelines.

Dr. Margarita López-Fernández, from the Department of Microbiology at the UGR (Spain), will be responsible for the data management within MicroBent project, and for creating a DMP and its subsequent updates, plus recording and updating the generated datasets.

5. DATA SECURITY

The Scientific Documentation Service of the UGR, located in the Library of the Hospital Real (Granada, Spain), coordinates the electronic management of the DIGIBUG repository. DIGIBUG incorporates a program for backup and preservation. Therefore, MicroBent datasets hosted in DIGIBUG will receive the same security treatment as the rest of the documents in this repository. All the responsibilities concerning data recovery and secure storage will go to the Scientific Documentation Service of the UGR, which oversees the storing of the datasets hosted in DIGIBUG.

6. ETHICAL ASPECTS

As stated in the Ethics Issues of the HORIZON-MSCA-2021-PF-01 grant agreement (101060228), there is no requirement for ethical review since MicroBent does not involve the use of human participants, human cells or tissues, personal data collection and/or processing, animals, potential for misuse of research results, or elements that may cause harm to the environment, animals or plants.